Extended Kalman Filter Design for Acetate Estimation in E. coli Cultures

L. Dewasme∗ G. Goffaux∗∗ A.-L. Hantson∗∗∗ A. Vande Wouwer∗

∗ Service d’Automatique, Université de Mons, Boulevard Dolez, 31, 7000 Mons, Belgium (e-mail: Laurent.Dewasme and Alain.VandeWouwer@umons.ac.be).
∗∗ Département de Génie Chimique, École Polytechnique de Montréal, C.P. 6079 Montréal, Québec, Canada (e-mail: Guillaume.Goffaux@polymtl.ca)
∗∗∗ Service de Chimie et Biochimie appliquées, Université de Mons, 56 Rue de l’Épargne, 7000 Mons, Belgium (e-mail: Anne-Lise.Hantson@umons.ac.be)

Abstract: Optimizing control of E. coli cultures depends on the availability of appropriate on-line sensors for the main culture components. A simple and efficient approach to maintain E. coli cultures in the neighborhood of the optimal operating conditions is to regulate the byproduct (i.e., acetate) concentration at a constant low level. Unfortunately, reliable acetate probes are currently not available on the market, and it is necessary to design a software sensor. In this work, observability conditions are first examined, and Extended Kalman Filters (EKF) are developed for various hardware sensor configurations, taking account of their reliability and cost. The filters are validated using experimental data collected on a lab-scale bioreactor.

Keywords: Observation, Software Sensors, Extended Kalman Filter, Fed-batch processes, Fermentation.

1. INTRODUCTION

Industrial recombinant protein production is usually achieved using fed-batch cultures of genetically modified yeast or bacterial strains. From an operational point of view, it is necessary to determine an optimal feeding strategy (i.e. the time evolution of the input flow rate to the fed-batch culture) in order to maximize the biomass productivity.

The main problem encountered comes from the metabolic changes in such strains in the case of a feeding excess. An "overflow metabolism", also called "short-term Crabtree effect" (Crabtree (1929), Deken (1966)), is a metabolic phenomenon that is induced when the rate of glycolysis (i.e., the oxidation of glucose) exceeds a critical value expressing the maximum capability of energy production by glucose oxidation, leading to a generally inhibiting byproduct formation from pyruvate. This phenomenon occurs for instance in S. cerevisiae cultures with aerobic ethanol formation, in P. pastoris with aerobic methanol formation, in E. coli cultures with aerobic acetate formation or in mammalian cell cultures with aerobic lactate formation. To avoid this undesirable effect, a closed-loop optimizing strategy is required, which can take various forms (Pomerleau (1990), Chen et al. (1995), Renard et al. (2006), Dewasme et al. (2009)).

Two main strategies are generally considered: a) controlling the substrate (generally glucose) concentration accurately at the critical level, i.e. the threshold separating the two metabolic regimes (with and without feeding excess); b) limiting the byproduct production by controlling its concentration at a very low level. Unfortunately, the sensitivity of currently available glucose probes is often insufficient to provide a correct measurement. Indeed, the critical level is very low (from $O(10^{-2})$ to $O(10^{-1})$). Consequently, the regulation of the byproduct concentration is preferred for obvious practical reasons. Nevertheless, even if ethanol and methanol are nowadays easily measurable, acetate and lactate probes are rare or nonexistent on the market. On the other hand, biomass, $pO_2$ probes and gas-analysers are now widely available, and can be the basis for the development of software sensors (see Bastin and Dochain (1990), Bogaerts and Vande Wouwer (2003)).

In Veloso et al. (2009), the performance of an Extended Kalman Filter (EKF) and an Asymptotic Observer (AO) is assessed in an application to E. coli fed-batch cultures. The observability of the model is studied for different combinations of the measured and estimated state variables. However, biomass is always considered as unmeasurable, limiting the number of possible combinations. Moreover, from all the assumed unmeasurable states, only the biomass concentration is adequately estimated while the performances of both observers are not satisfactory for the estimation of the substrate and acetate concentrations.

In Arndt et al. (2005), off-line glucose measurements by flow injection analysis (FIA) are used by a Kalman Filter to reconstruct the biomass signal and the glucose signal between 2 measurements. The estimated state vector is also augmented by the addition of the maximal growth rate. This Kalman Filter is coupled to a model-based PI controller to regulate the glucose concentration at 0.2 g/l. The set-point is arbitrarily chosen by the user and considered as efficient in the context of the application which aims at producing phytase. Moreover, a BL21(DE3) pPhyt109 E. coli strain is used. This strain is well-known for...
its property to produce small quantities of acetate, so that its accumulation only inhibits the biomass growth in the end of the culture.

In this work, a wild-type (B-11303, ATCC) strain is considered, which can produce larger amount of acetate, thus inhibiting the biomass growth in the early hours of the culture. This motivates the focus on the on-line estimation of the acetate concentration and, less importantly, the on-line estimation of the glucose concentration. In order to design the software sensors, biomass concentration is assumed to be measured on-line, so as the dissolved oxygen concentration and the outlet gas analysis.

This paper is organized as follows. The next section describes the process model. A global observability assessment is given in section 3 for various sensor configurations, together with a presentation of the EKF structure and some simulation results. Section 4 presents the available materials and methods, from the bioreactor set-up to the available probes and analyzers. Experimental results are presented in section 5. Conclusions are drawn in section 6.

2. A GENERIC MECHANISTIC MODEL

In this section, we first consider a generic mechanistic model that represents the culture of different strains presenting an overflow metabolism (yeasts, bacteria, animal cells, etc.).

In the following, three main catabolic reactions are considered:

Substrate oxidation: \( k_{S1} S + k_{O1} O \xrightarrow{r_{S}} k_{X1} X + k_{C1} C \)  

Substrate fermentation: \( k_{S2} S + k_{O2} O \xrightarrow{r_{F}} k_{X2} X + k_{P2} P + k_{C2} C \)

Byproduct oxidation: \( k_{P3} P + k_{O3} O \xrightarrow{r_{P}} k_{X3} X + k_{C3} C \)

Where \( X, S, P, O \) and \( C \) are, respectively, the concentration in the culture medium of biomass, substrate (typically glucose), byproduct (i.e. ethanol or methanol in yeast cultures, acetate in bacteria cultures or lactate in animal cells cultures), dissolved oxygen and carbon dioxide. \( k_{i} (i = 1, 2, 3, \xi = X, S, P, O, C) \) are the yield coefficients and \( r_{1}, r_{2} \) and \( r_{3} \) are the nonlinear specific consumption rates given by:

\[
\begin{align*}
  r_{1} &= \min \left( \frac{r_{S} r_{crit}}{k_{S1}}, 1 \right) \\
  r_{2} &= \max \left( 0, \frac{r_{O} r_{crit}}{k_{S2}} \right) \\
  r_{3} &= \max \left( 0, \frac{r_{P} r_{crit} - r_{S}}{k_{P3}} \right)
\end{align*}
\]

Note that these specific consumption rates are divided, for each reaction, by the corresponding yield coefficient (i.e. \( k_{S1} \) and \( k_{S2} \) for the main substrate, \( S \), in the first two reactions and \( k_{P3} \) for the consumed byproduct, \( P \), in the third reaction) in order to normalize with respect to the consumed source. For instance, 1 mole of substrate will produce \( k_{X1}/k_{S1} \) mole of biomass \( X \) with the first reaction and \( k_{X2}/k_{S2} \) mole of biomass \( X \) with the second reaction. The kinetic terms associated with the substrate consumption \( r_{S} \) and the critical substrate consumption \( r_{crit} \) (function of the cells oxidative or respiratory capacity \( r_{O} \)) are given by:

\[
\begin{align*}
  r_{S} &= \frac{\mu_{S}}{S + k_{S}} \\
  r_{crit} &= \frac{r_{O} k_{O1} O}{k_{O1} O + k_{O} k_{lip} + P}
\end{align*}
\]

In this work, a wild-type (B-11303, ATCC) strain is considered, which can produce larger amount of acetate, thus inhibiting the biomass growth in the early hours of the culture. This motivates the focus on the on-line estimation of the acetate concentration and, less importantly, the on-line estimation of the glucose concentration. In order to design the software sensors, biomass concentration is assumed to be measured on-line, so as the dissolved oxygen concentration and the outlet gas analysis.

These expressions take the classical form of Monod laws where \( \mu_{S} \) and \( \mu_{O} \) are the maximal values of specific growth rates and \( K_{S} \) and \( K_{O} \) are the saturation constants. Moreover, \( P \) inhibits the cells oxidative capacity and \( K_{lip} \) represents the inhibition constant.

The kinetic model (2) is based on Sonnleitner’s bottleneck assumption (Sonnleitner and Käppeli (1986)) which was applied to a yeast strain Saccharomyces cerevisiae (Figure 1). During a culture, the cells are likely to change their metabolism because of their limited oxidative capacity. When the substrate is in excess (concentration \( S > S_{crit} \) and the glucose consumption rate \( r_{S} > r_{Scrit} \)), the cells produce a byproduct \( P \) through the fermentative pathway, and the culture is said in respirative (R) regime. On the other hand, when the substrate becomes limiting (concentration \( S < S_{crit} \) and the glucose consumption rate \( r_{S} < r_{Scrit} \)), the available substrate (typically glucose), and possibly the byproduct \( P \) (as a substitute carbon source), if present in the culture medium, are oxidized. The culture is then said in respirative (R) regime.

Component-wise mass balances give the following differential equations:

\[
\begin{align*}
  \frac{dX}{dt} &= \left( k_{X1} r_{1} + k_{X2} r_{2} + k_{X3} r_{3} \right) X - DX \\
  \frac{dS}{dt} &= -(k_{S1} r_{1} + k_{S2} r_{2}) X + DS_{in} - DS \\
  \frac{dP}{dt} &= (k_{P2} r_{2} - k_{P3} r_{3}) X - DP \\
  \frac{dO}{dt} &= -(k_{O1} r_{1} + k_{O2} r_{2} + k_{O3} r_{3}) X - DO + OTR \\
  \frac{dC}{dt} &= (k_{C1} r_{1} + k_{C2} r_{2} + k_{C3} r_{3}) X - DC - CTR \\
  \frac{dV}{dt} &= F_{in}
\end{align*}
\]

where \( S_{in} \) is the substrate concentration in the feed, \( F_{in} \) is the inlet feed rate, \( V \) is the culture medium volume and \( D \) is the dilution rate \( (D = F_{in}/V) \). \( OTR \) and \( CTR \) represent respectively the oxygen transfer rate from the gas phase to the liquid phase and the carbon transfer rate from the liquid phase to the gas phase. Classical models of \( OTR \) and \( CTR \) are given by:

\[
\begin{align*}
  OTR &= k_{l} a_{O}(O_{sat} - O) \\
  CTR &= k_{l} a_{C}(C_{sat} - C)
\end{align*}
\]

where \( k_{l}a_{O} \) and \( k_{l}a_{C} \) are the volumetric transfer coefficients respectively of oxygen and carbon dioxide, and \( O_{sat} \) and \( C_{sat} \) are respectively the dissolved oxygen and carbon dioxide concentrations at saturation.
Several simplifications of (4) are possible, depending on the culture conditions. In controlled operation, the substrate concentration could be kept close to zero (in the following, this situation is denoted "regulation 1") leading to $P \approx 0$, $O \approx 0$ or alternatively (and equivalently), the byproduct concentration can be regulated at a low level ("regulation 2") leading to $P = 0$, $P << K_p$. In addition, ample oxygenation is provided and thus, $O = 0$ and $O >> K_O$. These assumptions will be considered in the following, when analyzing system observability and designing an observer.

3. DESIGN OF SOFTWARE SENSORS

3.1 Observability analysis

Observability can be assessed using canonical observability forms as introduced in Gauthier and Kupka (1994) and Zeitz (1984)

$$\forall i \in \{1, \ldots, q\}, x_i \in \mathbb{R}^n, n_1 \geq n_2 \geq \ldots \geq n_q, \sum_{i \leq q} n_i = n,$$

$$\dot{x} = \begin{cases} x_1 = f_1(x_1, x_2) \\ x_2 = f_2(x_1, x_2, x_3) \\ \vdots \\ x_{n_i - 1} = f_{q_i - 1}(x_1, \ldots, x_q) \\ x_i = f_{q_i}(x_1, \ldots, x_q) \\ y = x_1 = [x_{1,1} x_{1,2} \ldots x_{1,n_1}]' \end{cases} \tag{6}$$

where $x$ is the state vector, $y$ the vector of measured states, $f_i$ a partition of the nonlinear state equations, $q$ the number of partitions. (6) is also called a Lower Hessenberg System, i.e., a system where $\frac{\partial f_i}{\partial x_j} = 0$ with $j > i + 1$.

To assess if the system is observable, one first checks if the bioprocess model can be put in the form of (6) by defining an appropriate partition, and then the following condition is evaluated:

$$\text{rank} \left[ \frac{\partial f_i}{\partial x_{i+1}} \right] = n_{i+1} \quad \forall i \in \{1, \ldots, q - 1\} \tag{7}$$

(7) simply translates the fact that a partition of states $x_{i+1}$ is only observable if any perturbation of these states propagates to partition $x_i$.

Model (4) can be put in the generic form:

$$\dot{x} = \begin{cases} \dot{X} = f_1(X, S, P, O) \\ \dot{O} = f_2(X, S, P, O) \\ \dot{S} = f_2(X, S) \\ \dot{P} = f_3(X, S, P, O) \end{cases} \tag{8}$$

Note that the carbon dioxide differential equation is not considered as $C$ is usually not measured, does not need to be estimated and does not influence the other states. More interestingly, the substrate concentration (i.e., more exactly, its variation) is only a function of itself and the biomass concentration. This is due to the particular forms of the reaction rates (2) leading to a simplification of the kinetic term, reduced to $r_S X$. Following the condition on the $n_i$ in (6), it appears that the only partition satisfying conditions (6) and (7) and allowing the estimation of $S$ and $P$ is $y = x_1 = [X]$ and $x_2 = [S P]'$. Indeed, the system becomes:

$$\begin{align*}
\dot{x}_1 &= [X] \\
\dot{x}_2 &= [S P]' = f_1(x_1, x_2) \\
y &= x_1 = [X]' \tag{9}
\end{align*}$$

and

$$\begin{align*}
\text{rank} \frac{\partial f_1}{\partial x_{i+1}} &= \text{rank} \begin{bmatrix} \frac{\partial X}{\partial O} & \frac{\partial X}{\partial P} \\ \frac{\partial O}{\partial S} & \frac{\partial O}{\partial P} \end{bmatrix} \\
&= \begin{bmatrix} X \mu S_k & k_o X \mu S_k (S + K_S)^2 \\ -X \mu S_k & k_o X \mu S_k (S + K_S)^2 \end{bmatrix} \\
&= n_2 = 2 \tag{10}
\end{align*}$$

As mentioned at the end of the previous section, several simplifying assumptions hold in controlled conditions, and in particular for regulation 1: $S \approx 0$, $S \approx 0$, $O = 0$ and $O >> K_O$. System (8) can therefore be rewritten as follows:

$$\begin{align*}
\dot{x} &= \begin{cases} \dot{X} = f_1(X, P) \\
\dot{P} = f_2(X, P) \\
y = X \tag{11}
\end{cases}
\end{align*}$$

In the particular case of (11), $n_i = 1$ $\forall i$ and condition (7) reduces to $\frac{\partial f_i}{\partial x_{i+1}} \neq 0$ which is verified if $X, O \neq 0$ (Note that in order to achieve an aerobic culture, the biomass and the oxygen concentrations can never vanish). It is also possible to simplify (8) according to regulation 2: $P = 0$ and $K_p >> P$, together with $O = 0$ and $O >> K_O$:

$$\begin{align*}
\dot{x} &= \begin{cases} \dot{X} = f_1(X, S) \\
\dot{S} = f_2(X, S) \\
y = X \tag{12}
\end{cases}
\end{align*}$$

and $\frac{\partial S}{\partial P} \neq 0$ if $X \neq 0$. Combining (11) and (12), it appears that the biomass measurement could be sufficient to reconstruct the byproduct and the substrate concentrations as long as those concentrations remain at respective sufficiently low levels (i.e., $S < O(10^{-1} \text{ g/l})$ and $0 \text{ g/l} < A < 1 \text{ g/l}$).

System observability can be assessed for various sensor configurations using the bioprocess model (8) or a simplified version of it based on the assumptions detailed in section 2. The best sensor configuration will depend on the previous observability analysis, economic considerations ($pO_2$ probes are generally cheaper than a gas analyser, which is cheaper than biomass and substrate probes), and practical observability as assessed from the performance of a software sensor running under realistic conditions of measurement noise. To this end, we first introduce the extended Kalman Filter (EKF).

3.2 The Extended Kalman Filter

The Kalman filter (Gelb (1974)), which is by far the most popular state estimation technique used for bioprocess monitoring, is an exponential observer that minimizes the variance of the estimation error. The EKF is based on a first-order linearization
of the process model along the estimated trajectory. In the context of bioprocess applications, the concentrations of the main species are measured at discrete times only and with relatively low sampling frequencies (the measurements are sometimes collected at different rates, i.e., resulting in an asynchronous measurement configuration). Therefore, an interesting EKF formulation is the continuous-discrete where continuous estimations are provided from discrete measurements. The algorithm proceeds in two steps: a prediction step (corresponding to the time period between two measurement times) and a correction step occurring each time a new measurement is available.

The first step (prediction between \( t_k \) and \( t_{k+1} \)) corresponds to:

\[
\frac{d\xi(t)}{dt} = K\phi(\xi(t), t) - D\xi(t) + D\xi_{in} ; \quad \xi(t_k) = \xi(t_k^+), \quad t_k \leq t < t_{k+1}
\]  

(13a)

\[
\frac{dC(t)}{dt} = A(\xi(t))C(t) + C(t)A(\xi(t))^T + R_\Omega ; \quad C(t_k) = C(t_k^+),
\]  

(13b)

The correction step occurring at time \( t_{k+1} \) corresponds to:

\[
\Omega(\xi(t_{k+1})) = C(t_{k+1})L^T [LC(t_{k+1})L^T + R_k(t_{k+1})]\^{-1}
\]  

(14a)

\[
\xi(t_{k+1}^-) = \xi(t_{k+1}^+) + \Omega(\xi(t_{k+1}))(y(t_{k+1}) - L\xi(t_{k+1}^-))
\]  

(14b)

\[
C(t_{k+1}^-) = C(t_{k+1}^-) - \Omega(\xi(t_{k+1}^-))LC(t_{k+1}^-)
\]  

(14c)

where \( \xi \) is the vector of concentrations of the macroscopic components, \( K \) the pseudo-stoichiometric matrix, \( \phi \) the vector of reaction rates, \( L \) the measurements matrix, \( \Omega \) the correction gain, \( C \) the covariance matrix of the state errors, \( R_k \) and \( R_\Omega \) the covariance matrices of respectively the measurement and model noises and \( t_{k+1}^- \) and \( t_{k+1}^+ \) the moments characterizing respectively the values before and after correction.

This filter is first applied to the bacterial culture model (4) with published parameter values, e.g., the parameter values from Rocha (2003), using various sensor configurations, and realistic measurement noise. The noise standard deviation is chosen as \( \sigma = 0.0001 \) g/l for the dissolved oxygen measurement, \( \sigma = 0.5 \) g/l for the biomass measurement, \( \sigma = 0.05 \) g/l for the glucose measurement and \( \sigma = 0.01 \) g/l/s for the OTR measurement.

Table 1 summarizes the results of this analysis. The first column details the different measurement combinations, the second column specifies the model (the full model (4) or the simplified model obtained using the assumptions described in section 2 for which canonical observability is verified and the last 2 columns give a quantitative statement regarding the practical observability of the byproduct (P for acetate) and the substrate.

This estimation error is assessed as \( \varepsilon = \frac{1}{N} \sum_{i=1}^{N} (\xi_i - \tilde{\xi}_i)^2 \) where \( N \) is the number of samples and \( \tilde{\xi}_i \) is the vector of the estimated states.

The cheapest way to estimate the byproduct concentration is to measure the dissolved oxygen concentration only, as shown in Table 1 and Figure 2. However, in realistic conditions, the dissolved oxygen is regulated by the stirrer speed or the air flow, requiring an additional measurement from the gas analyzer (i.e., the OTR) to recover information on the oxygen consumption. The byproduct and substrate estimations from the oxygen measurement can qualitatively be assessed as correct (see Figure 2) and not necessarily improve if the biomass concentration is also available, as shown by Figure 3 where biomass and dissolved oxygen concentrations are both measured.

An observability analysis using the simplifying assumptions of model (4) discussed in section 2 shows that the biomass measurement could be sufficient to estimate the byproduct and the substrate concentrations as long as these concentrations remain sufficiently low. Indeed, Figure 4 shows that acetate and substrate are both well estimated.

The combinations using the substrate concentration measurement are not considered in the following as it is the most expensive measurement device and does not provide any improvement, as compared to the observer using the biomass and/or the dissolved oxygen measurements.

### 4. MATERIALS AND METHODS

#### 4.1 Bioreactor setup

Fermentations of a wild-type strain (B) of \( E. \) coli (B-11303, ATCC) are performed in a 5 l compact laboratory scale bioreactor (Biostat B + - Sartorius). pH is regulated at 7 by a PI controller acting on the injection of NaOH 2 M (base) and H\(_2\)PO\(_4\) 0.5 M (acid). Oxygenation conditions are regulated through dissolved oxygen control by a 2-stage cascade controller acting first on the stirrer speed (from 200 to 1500 RPM) and, after stirrer speed saturation, on the aeration system supplying air.
measurement, the samples are buffered in the mobile phase
(4
HPLC
(Off-line acetate measurements can be achieved by an
measurement technique is appropriate in batch operating con-
dditions where glucose concentrations are sufficiently high but
delicate to use in fed-batch conditions where the glucose
concentration is too small to allow an accurate measurement.

4.2 Glucose measurement

Off-line glucose measurements are performed by a Process
Trace enzymatic system. Note that on-line measurements can also be
achieved through the same system using a dialysis probe or a microfiltration probe. However, the highest sensitivity of available measurement kits is 0.01 g/l. Unfortunately, the level of the critical glucose concentration is, of course, a priori unknown but is in the range of this highest sensitivity. This measurement technique is appropriate in batch operating conditions where glucose concentrations are sufficiently high but more delicate to use in fed-batch conditions where the glucose concentration is too small to allow an accurate measurement.

4.3 Off-line acetate measurement

Off-line acetate measurements can be achieved by an Alliance
HPLC (Waters, USA) using a 3 µm Atlantis C18 column
(4.6x150 mm, Waters, USA) at 30°C and a UV detector 486
(Waters, USA) set at 210 nm in isocratic mode with a
NaH2PO4 20 mM solution as mobile phase (using a flow rate
of 0.5 ml/min). To ensure reproducibility and stability of the
measurement, the samples are buffered in the mobile phase
prior to injection. For comparison purposes, an acetic acid kit purchased from Megazyme (Ireland) is also used.

4.4 On-line gas measurement

Molar fractions of O2 and CO2 in the outlet gas are available
thanks to gas analysis using a DUET — gas analyser from
System — C — Industry. As explained in Rocha (2003), OTR and
CTR can be measured from the knowledge of molar fractions of
O2, CO2 and N2 in the outlet gas, the volume and the inlet
air flow. In the next experimental applications, the inlet air flow is
chosen equal to 1 l/min and O2in and CO2in are measured in
the outlet gas before starting the culture.

5. EXPERIMENTAL RESULTS

5.1 Model identification

Starting from the general model (4), a simplified reaction
scheme taking only the respiro-fermentative pathway into ac-
count (i.e., oxidation and fermentation of glucose) and consider-
ing that oxygen conditions are not limiting the cells growth is
derived as follows:

Substrate oxidation :

\[ k_{S1}S + k_{O1}O \rightarrow k_{r1}S \rightarrow k_{X1}X \]  (15a)

Substrate fermentation :

\[ k_{2}S + k_{2}O \rightarrow k_{r2}S \rightarrow k_{X2}X + k_{P2}P \]  (15b)

The specific growth rates are:

\[ r_1 = \frac{\min (r_S, r_{S\rightarrow P})}{k_{S1}} \]  (16)
\[ r_2 = \frac{\max (0, r_S - r_{S\rightarrow P})}{k_{S2}} \]  (17)

In the proposed situation, as the oxygen is assumed not to
be limiting, \( K_O \) is neglected in (3b) so that
\[ r_{S\rightarrow P} = \frac{\mu_O}{k_{O1}} \approx \frac{\mu_O}{k_{O1}} \frac{K_{ip}}{k_{O1} K_{ip} + P} \]

The kinetic terms associated with the substrate consumption \( r_S \) and the critical substrate consumption \( r_{S\rightarrow P} \) are:

\[ r_S = \frac{\mu_S}{S + K_S} \]  (18a)
\[ r_{S\rightarrow P} = \frac{r_O}{k_{O1}} \approx \frac{\mu_O}{k_{O1}} \frac{K_{ip}}{k_{O1} K_{ip} + P} \]  (18b)

Therefore, assuming the existence of the fermentation reaction and taking (18) into account, the specific growth rates take the following forms:

\[ r_1 = \frac{r_{S\rightarrow P}}{k_{S1}} \]  (19)
\[ r_2 = \frac{r_S - r_{S\rightarrow P}}{k_{S2}} \]  (20)

Finally, normalizing the yield coefficients with respect to the
substrate (\( k_{S1} = k_{S2} = 1 \)), the simplified reaction scheme gives the following component-wise mass balances differential equation system:

\[ \frac{dX}{dt} = (k_{X1}r_{S\rightarrow P} + k_{X2}(r_S - r_{S\rightarrow P}))X - DX \]  (21a)
\[ \frac{dS}{dt} = -r_SX - D(S - S_{in}) \]  (21b)
\[ \frac{dP}{dt} = k_{P2}(r_S - r_{S\rightarrow P})X - DP \]  (21c)

The following parameter values are obtained after identification:
\[ k_{X1} = 1.742, k_{X2} = 0.305, k_{P2} = 1.196, k_{O1} = 0.974, k_{O2} = 0.197, \mu_S = 0.72, K_S = 0.05, \mu_O = 0.0935 \text{ and } K_{ip} = 4. \]
5.2 Acetate and glucose estimations using an EKF

The EKF presented in 3.2 is used in order to estimate the acetate and glucose concentrations on the basis of on-line biomass concentration measurements from a Sartorius™ turbidimetric probe and the identified model (21) with identified parameter values from section 5.1.

The biomass concentration signal is generally affected by white noise with a standard deviation around 0.5 g/l. Consequently, the covariance matrix of the measurement noise is selected as $R_e = 25 \times 10^{-3}$. Moreover, the confidence degree in the model is chosen through the covariance matrix of the model noise, which in this case is $R_n = \begin{bmatrix} 10^{-2} & 0 & 0 \\ 0 & 10^{-2} & 0 \\ 0 & 0 & 10^{-2} \end{bmatrix}$. Note that biomass on-line measurements present a very high sampling (the sampling period is 5 s) so that they appear as continuous in the next results.

For the shown experiment, the observer obviously performs better than the identified model (which, as the biomass grows, drifts away from the turbidimetric on-line measurements) testifying of a certain level of robustness against model errors thanks to the biomass measurement. Indeed, glucose and acetate estimations fit very well with the off-line measurements. However, note that the confidence interval increases with time, indicating a decreasing accuracy of the acetate estimation. During the fed-batch phase (i.e., when the glucose concentration becomes close to 0 g/l), off-line measurements also lack accuracy and always indicate the sensitivity level (0.1 g/l) so that one can imagine that the estimation of the EKF around 0.01 g/l is more realistic as the confidence interval is tight.

6. CONCLUSION

The availability of biomass probes offers new possibilities for estimating the acetate and glucose concentrations in fed-batch cultures of E. coli. Theoretical and practical observability analyses show that the best sensor configuration consists of a biomass probe together with the measurement of dissolved oxygen and off-gas analysis. However, the sole measurement of biomass already provides satisfactory results, when the underlying bioprocess model is accurate enough (i.e. has been carefully identified from experimental data). Future research is dedicated to the experimental assessment of the robustness of the estimation scheme, and the possibility of adapting a few observer parameters using the extra information contained in the dissolved oxygen concentration and off-gas analysis. Indeed, the main perspective is to insert a robust version of the presented filter in a control scheme dedicated to cell growth optimization.

REFERENCES


